

Bounds on the Reconstruction of Sparse Signal Ensembles from Distributed Measurements

Marco F. Duarte, *Member, IEEE*, Michael B. Wakin, *Member, IEEE*,
Dror Baron, *Senior Member, IEEE*, Shriram Sarvotham, *Member, IEEE*,
and Richard G. Baraniuk, *Fellow, IEEE*

This paper is dedicated to the memory of Hyeokho Choi, our colleague, mentor, and friend.

Abstract

In compressive sensing, a small collection of linear projections of a sparse signal contains enough information to permit signal recovery. Distributed compressive sensing (DCS) extends this framework, allowing a correlated ensemble of sparse signals to be jointly recovered from a collection of separately acquired compressive measurements. In this paper, we introduce an ensemble sparsity model for capturing the intra- and inter-signal correlations within a collection of sparse signals. For strictly sparse signals obeying an ensemble sparsity model, we characterize the fundamental number of noiseless measurements that each sensor must collect to ensure that the signals are jointly recoverable. Our analysis is based on a novel bipartite graph representation that links the sparse signal coefficients with the measurements obtained for each signal.

Index Terms

Compressive sensing, random projections, signal ensembles, sparsity.

Manuscript submitted February 14, 2011. This work was supported by the grants NSF CCF-0431150, NSF CCF-0926127, DARPA HR0011-08-1-0078, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936, ONR N00014-08-1-1112, AFOSR FA9550-09-1-0432, AFOSR FA9550-09-1-0465, ARO MURI W911NF-07-1-0185, ARO MURI W911NF-09-1-0383, and the Texas Instruments Leadership University Program. MFD was also supported by NSF Supplemental Funding DMS-0439872 to UCLA-IPAM, P.I. R. Caflisch. DB was also supported by the Israel Science Foundation while at the Technion – Israel Institute of Technology.

M. F. Duarte is with the Department of Computer Science, Duke University, Durham, NC, 27707 USA. E-mail: marco.duarte@duke.edu

M. B. Wakin is with the Division of Engineering, Colorado School of Mines, Golden, CO, 80401 USA. E-mail: mwakin@mines.edu

D. Baron is with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, 27695 USA. E-mail: barondror@ncsu.edu

S. Sarvotham is with Halliburton Energy Services, Houston, TX, 77032 USA. E-mail: shri@rice.edu

R. G. Baraniuk is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX, 77005 USA. E-mail: richb@rice.edu

I. INTRODUCTION

A unique framework for signal sensing and compression has recently developed under the rubric of *compressive sensing* (CS). CS builds on the work of Candès, Romberg, and Tao [1] and Donoho [2], who showed that if a signal $x \in \mathbb{R}^N$ can be expressed as a sparse superposition of just $K < N$ elements from some dictionary, then it can be recovered from a small number of linear measurements $y = \Phi x$, where Φ is a measurement matrix of size $M \times N$, and $M < N$. One intriguing aspect of CS is that randomly chosen measurement matrices can be remarkably effective for nonadaptively capturing the information in sparse signals. In fact, if x is a fixed K -sparse signal and just $M = K + 1$ random measurements are collected via a matrix Φ with independent and identically distributed (i.i.d.) Gaussian entries, then with probability one x is the unique K -sparse solution to $y = \Phi x$ [3]. While there are no tractable recovery algorithms that guarantee recovery when so few measurements are collected, there do exist a variety of practical and provably effective algorithms [1, 2, 4, 5] that work when $M = O(K \log N)$.

The current CS theory has been designed mainly to facilitate the sensing and recovery of a single signal $x \in \mathbb{R}^N$. It is natural to ask whether CS could help alleviate the burdens of acquiring and processing high-dimensional data in applications involving multiple sensors. Some work to date has answered this question in the affirmative. For example, if the entries of an unknown vector $x \in \mathbb{R}^N$ are spread among a field of sensors (e.g., if x represents a concatenation of the ambient temperatures recorded by N sensors at a single instant), then certain protocols have been proposed for efficiently computing $y = \Phi x$ through proper coordination of the sensors [6–9]. Given y , standard CS recovery schemes can then be used to recover x using a model for its sparse structure.

It is interesting, however, to consider cases where each sensor observes not a single scalar value but rather a longer vector. For example, consider an ensemble of signals $x_1, x_2, \dots, x_J \in \mathbb{R}^N$ observed by a collection of J sensors, where each sensor $j \in \{1, 2, \dots, J\}$ observes only signal x_j (e.g., x_j might represent a time series of N temperature recordings at sensor position j). In such a scenario, one could employ CS on a sensor-by-sensor basis, recording random measurements $y_j = \Phi_j x_j$ of each signal, and then reconstructing each signal x_j from the measurements y_j . Such an approach would exploit *intra*-signal correlations (manifested in a sparse model for each x_j), but would not exploit any *inter*-signal correlations that may exist among the signals x_j .

Motivated by this observation, we have proposed a framework known as *distributed compressive sensing* (DCS) that allows the exploitation of *both* intra- and inter-signal correlation structures.¹ In a typical DCS scenario, each sensor separately collects measurements $y_j = \Phi_j x_j$ as described above, but these measurements are then transmitted to a single collection point (a single “decoder”) where the ensemble of signals is reconstructed *jointly* using a model that characterizes correlations among the sparse signals. By exploiting the inter-signal correlations, DCS allows the overall measurement burden to be shared among the J sensors; in other words, the signal ensemble can be reconstructed jointly from significantly fewer measurements than would be required if each signal were reconstructed individually. Although we do not go into the details here, one can make interesting connections between DCS and the Slepian-Wolf framework for distributed source coding, in which correlated random sources

¹Our prior work in DCS is contained in two technical reports [3, 10] and several conference publications [11–15].

can each be encoded below their nominal entropy rate if they are decoded jointly [3, 11, 16, 17].

As mentioned above, any DCS decoder must rely on a correlation model that describes the anticipated structure within and among the signals in the ensemble. There are many conceivable ways in which correlations can be described among a collection of sparse signals. We have previously proposed [3, 11–14] several models for capturing such correlations and studied each model in isolation, developing a variety of practical reconstruction algorithms and theoretical arguments customized to the nuances of each model. The goal of this paper is to develop a broader, general purpose framework for quantifying the sparsity of an ensemble of correlated signals. In Section II, we introduce a factored representation of the signal ensemble that decouples its location information from its value information: a single vector encodes the values of all nonzero signal entries, while a binary matrix maps these values to the appropriate locations in the ensemble. We term the resulting models *ensemble sparsity models* (ESMs). ESMs are natural and flexible extensions of single-signal sparse models; in fact, the ensemble correlation model proposed in [18] and all of our previously proposed models fit into the ESM framework as special cases.

The bulk of this paper (Section III, Section IV, and several supporting appendices) is dedicated to answering a fundamental question regarding the use of ESMs for DCS: how many measurements must each sensor collect to ensure that a particular signal ensemble is recoverable? Not surprisingly, this question is much more difficult to answer in the multi-signal case than in the single-signal case. For this reason, we focus not on tractable recovery algorithms or robustness issues but rather on the foundational limits governing how the measurements can be amortized across the sensors while preserving the information required to uniquely identify the sparse signal ensemble. To study these issues, we introduce a bipartite graph representation generated from the ESM that reflects, for each measurement, the sparse signal entries on which it depends. Our bounds relate intimately to the structure of this graph. While our previous work in DCS has helped inspire a number of algorithms for recovery of real-world signal ensembles [19–21], we believe that the results in this paper and the analytical framework that we introduce will help establish a solid foundation for the future development of DCS theory.

II. ENSEMBLE SPARSITY SIGNAL MODELS

In this section, we propose a general framework to quantify the sparsity of an ensemble of correlated signals. Our approach is based on a factored representation of the signal ensemble that decouples its location information from its value information. Later, in Section III, we explain how the framework can be used in the joint recovery of sparse signals from compressive measurements, and we describe how such measurements can be allocated among the sensors.

A. Notation and Definitions

We use the following notation for signal ensembles. Let $\Lambda := \{1, 2, \dots, J\}$ index the J signals in the ensemble. For a subset $\Gamma \subseteq \Lambda$, we define $\Gamma^C := \Lambda \setminus \Gamma$. Denote the signals in the ensemble by x_j , with $j \in \Lambda$. We assume that each signal $x_j \in \mathbb{R}^N$, and we let

$$X = [x_1^T \ x_2^T \ \dots \ x_J^T]^T \in \mathbb{R}^{JN}$$

denote the concatenation of the signals. For a given vector v , we use $v(n)$ to denote the n^{th} entry of v , and we use the ℓ_0 “norm” $\|v\|_0$ to denote the number of nonzero entries in v . Conventionally, $\|v\|_0$ is referred to as the *sparsity* of the vector v ;² we elaborate on this point below and discuss natural extensions of the concept of sparsity to multi-signal ensembles. Finally, we will also refer to the following definition.

Definition 1: For any nonnegative integers $L_1 \leq L_2$, an $L_2 \times L_1$ *identity submatrix* is a matrix constructed by selecting L_1 columns from the $L_2 \times L_2$ identity matrix $\mathbf{I}_{L_2 \times L_2}$; the selected columns need not be adjacent, but their order is preserved from left to right.

B. Sparse Modeling for a Single Signal

To motivate the use of a factored representation for modeling sparsity, we begin by considering the structure of a single sparse signal $x \in \mathbb{R}^N$ that has $K \leq N$ nonzero entries. We note that the degrees of freedom in such a signal are captured in the K locations where the nonzero coefficients occur and in the K nonzero values at these locations. It is possible to decouple the location information from the value information by writing $x = P\theta$, where $\theta \in \mathbb{R}^K$ contains only the nonzero entries of x , and where P is an $N \times K$ identity submatrix that includes x in its column span. Any K -sparse signal can be written in this manner.

In light of the above, to model the set of all possible sparse signals, define \mathcal{P} to be the set of all identity submatrices of all possible sizes $N \times K'$, with $1 \leq K' \leq N$. We refer to \mathcal{P} as a *sparsity model*, because the concept of sparsity can in fact be defined within the context of this model. To be specific, given an arbitrary signal $x \in \mathbb{R}^N$, one can consider all possible factorizations $x = P\theta$ with $P \in \mathcal{P}$. Among these factorizations, the dimensionality of the unique smallest representation θ equals the *sparsity level* of the signal x ; in other words, we will have $\dim(\theta) = \|x\|_0$.

C. Sparse Modeling for a Signal Ensemble

We generalize the formulation of Section II-B to the signal ensemble case by considering factorizations of the form $X = P\Theta$, where $X \in \mathbb{R}^{JN}$ represents the entire signal ensemble as defined above, P is a matrix of size $JN \times Q$ for some integer Q , and $\Theta \in \mathbb{R}^Q$. In any such factorization, we refer to P and Θ as the *location matrix* and *value vector*, respectively.

Definition 2: An *ensemble sparsity model* (ESM) is a set \mathcal{P} of admissible location matrices P ; the number of columns among the $P \in \mathcal{P}$ may vary, but each has JN rows.

As we discuss further below, there are a number of natural choices for what should constitute a valid location matrix P , and consequently, there are a number of possible ESMs that could be used to describe the correlations among sparse signals in an ensemble.

For a fixed ESM, not every matrix $P \in \mathcal{P}$ can be used to generate a given signal ensemble X .

²We consider for the sake of illustration—but without loss of generality—signals that are sparse in the canonical basis. All of our analysis can be easily extended to signals that are sparse in any orthonormal basis.

Definition 3: For a given ensemble X and ESM \mathcal{P} , the set of *feasible location matrices* is

$$\mathcal{P}_F(X) := \{P \in \mathcal{P} \text{ s.t. } X \in \text{colspan}(P)\},$$

where $\text{colspan}(P)$ denotes the column span of P .

Note that $\mathcal{P}_F(X) \subseteq \mathcal{P}$.

Definition 4: In the context of an ESM \mathcal{P} , the *ensemble sparsity level* of a signal ensemble X is

$$D = D(X, \mathcal{P}) := \min_{P \in \mathcal{P}_F(X)} \dim(\text{colspan}(P)).$$

When P is full-rank, the dimension of its column span is equal to its number of columns; we will expand on this property in Section II-E. For many ESMs, we may expect to have $D < \sum_{j \in \Lambda} \|x_j\|_0$.

D. Common/Innovation Location Matrices

There are a number of natural choices for the location matrices P that could be considered for sparse modeling of a signal ensemble. In this paper (as we studied earlier in [3]), we are interested in the types of multi-signal correlations that arise when a number of sensors observe a common phenomenon (which may have a sparse description) and each of those same sensors observes a local innovation (each of which may also have a sparse description). To support the analysis of such scenarios, we restrict our attention in this paper to location matrices P of the form

$$P = \begin{bmatrix} P_C & P_1 & \mathbf{0} & \dots & \mathbf{0} \\ P_C & \mathbf{0} & P_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ P_C & \mathbf{0} & \mathbf{0} & \dots & P_J \end{bmatrix}, \quad (1)$$

where P_C and each P_j , $j \in \Lambda$, are identity submatrices with N rows, and where each $\mathbf{0}$ denotes a matrix of appropriate size with all entries equal to 0. For a given matrix P of this form, let $K_C(P)$ denote the number of columns of the element P_C contained in P , and for each $j \in \Lambda$, let $K_j(P)$ denote the number of columns of P_j .

Let us explain why such location matrices are conducive to the analysis of signals sharing the common/innovation structure mentioned above. When a signal ensemble $X \in \mathbb{R}^{JN}$ is expressed as $X = P\Theta$ for some P of the form (1), we may partition Θ into the corresponding components

$$\Theta = [\theta_C^T \ \theta_1^T \ \theta_2^T \ \dots \ \theta_J^T]^T,$$

where $\theta_C \in \mathbb{R}^{K_C(P)}$ and each $\theta_j \in \mathbb{R}^{K_j(P)}$. Then, letting

$$z_C := P_C \theta_C \quad \text{and} \quad z_j := P_j \theta_j \text{ for each } j \in \Lambda, \quad (2)$$

we can write each signal in the ensemble as

$$x_j = z_C + z_j,$$

where the *common component* z_C has sparsity $K_C(P)$ and is present in each signal and the *innovation components* z_1, z_2, \dots, z_J have sparsities $K_1(P), K_2(P), \dots, K_J(P)$, respectively, and are unique to the individual signals.

Example 1: Consider $J = 2$ signals of dimension $N = 4$ each, specifically $x_1 = [3 \ 1 \ 0 \ 0]^T$ and $x_2 = [1 \ 1 \ 0 \ 0]^T$. Different choices of P can account for the common structure in x_1 and x_2 in different ways. For example, we could take

$$P_C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad P_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \text{and} \quad P_2 = [\], \quad (3)$$

in which case we can write $X = P\Theta$ by taking $\Theta = [1 \ 1 \ 2]^T$. Under this choice of P , we have $z_C = [1 \ 1 \ 0 \ 0]^T$, $z_1 = [2 \ 0 \ 0 \ 0]^T$ and $z_2 = [0 \ 0 \ 0 \ 0]^T$, and the sparsity levels for the respective components are $K_C(P) = 2$, $K_1(P) = 1$, and $K_2(P) = 0$. Alternatively, we could take

$$\tilde{P}_C = P_C, \quad \tilde{P}_1 = P_2, \quad \text{and} \quad \tilde{P}_2 = P_1, \quad (4)$$

in which case we can write $X = \tilde{P}\tilde{\Theta}$ by taking $\tilde{\Theta} = [3 \ 1 \ -2]^T$. Under this choice of \tilde{P} , we have $\tilde{z}_C = [3 \ 1 \ 0 \ 0]^T$, $\tilde{z}_1 = [0 \ 0 \ 0 \ 0]^T$ and $\tilde{z}_2 = [-2 \ 0 \ 0 \ 0]^T$, and the sparsity levels for the respective components are $K_C(\tilde{P}) = 2$, $K_1(\tilde{P}) = 0$, and $K_2(\tilde{P}) = 1$.

E. Common/Innovation ESMs

In this paper, we restrict our attention to ESMs that are populated only with a selection of the common/innovation location matrices described in Section II-D.

Definition 5: An ESM \mathcal{P} is called a *common/innovation ESM* if every $P \in \mathcal{P}$ has the form (1) and is full-rank.

The requirement that each $P \in \mathcal{P}$ have full rank forbids any P for which P_C and all $\{P_j\}_{j \in \Lambda}$ have one or more columns in common; it is natural to omit such matrices, since a full-rank matrix of the form (1) could always be constructed with equivalent column span by removing each shared column from P_C or any one of the P_j .

Depending on the type of structure one wishes to characterize within an ensemble, a common/innovation ESM \mathcal{P} could be populated in various ways. For example:

- One could allow \mathcal{P} to contain *all* full-rank matrices P of the form (1). This invokes a sparse model for both the common and innovation components.
- Or, one could consider only full-rank matrices P of the form (1) where $P_C = \mathbf{I}_{N \times N}$. This removes the assumption that the common component is sparse.
- Alternatively, one could consider only full-rank matrices P of the form (1) where $P_C = [\]$ and where $P_1 = P_2 = \dots = P_J$. This model assumes that no common component is present, but that the innovation components all share the same sparse support.
- Finally, one could consider only full-rank matrices P of the form (1) where $P_C = [\]$ and where all of the matrices P_j share some minimum number of columns in common. This model assumes that all innovations components share some support indices in common.

We have previously studied each of the first three cases above [3, 11–14], proposing a variety of practical reconstruction algorithms and theoretical arguments customized to the nuances of each model. Later, the fourth case above was proposed and studied in [18]. In this paper, however, we present a unified formulation, treating each model as a special case of the more general common/innovation ESM framework. Consequently, the theoretical foundation that we develop starting in Section III is agnostic to the choice of *which* matrices P of the form (1) are chosen to populate a given ESM \mathcal{P} under consideration, and therefore our results apply to all of the cases in [3, 11–14, 18].

III. DISTRIBUTED MEASUREMENT BOUNDS

In this section, we present our main results concerning the measurement and reconstruction of signal ensembles in the context of common/innovation ESMs.

A. Distributed Measurements

We focus on the situation where distributed measurements of the signals in an ensemble $X \in \mathbb{R}^{JN}$ are collected. More precisely, for each $j \in \Lambda$, let Φ_j denote a measurement matrix of size $M_j \times N$, and let $y_j = \Phi_j x_j$ represent the measurements collected of component signal x_j . When appropriate below, we make explicit an assumption that the matrices Φ_j are drawn randomly with i.i.d. Gaussian entries, though other random distributions could also be considered.

Suppose that the collection of measurements

$$Y = [y_1^T \ y_2^T \ \cdots \ y_J^T]^T \in \mathbb{R}^{\sum_{j=1}^J M_j}$$

is transmitted to some central node for reconstruction. Defining

$$\Phi = \begin{bmatrix} \Phi_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Phi_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Phi_J \end{bmatrix} \in \mathbb{R}^{(\sum_{j=1}^J M_j) \times JN},$$

we may write $Y = \Phi X$. In the context of a common/innovation ESM \mathcal{P} , we are interested in characterizing the requisite numbers of measurements M_1, M_2, \dots, M_J that will permit the decoder to perfectly reconstruct the ensemble X from Y .

B. Reconstruction of a Value Vector

Let us begin by considering the case where the decoder has knowledge of some full-rank location matrix $P \in \mathcal{P}_F(X)$. In this case, perfect reconstruction of the ensemble X is possible if the decoder can identify the unique value vector Θ such that $X = P\Theta$.

To understand when perfect reconstruction may be possible, note that for any Θ such that $X = P\Theta$, we can write

$$Y = \Phi X = \Phi P \Theta = \underbrace{\begin{bmatrix} \Phi_1 P_C & \Phi_1 P_1 & \mathbf{0} & \dots & \mathbf{0} \\ \Phi_2 P_C & \mathbf{0} & \Phi_2 P_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Phi_J P_C & \mathbf{0} & \mathbf{0} & \dots & \Phi_J P_J \end{bmatrix}}_{\Gamma} \underbrace{\begin{bmatrix} \theta_C \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_J \end{bmatrix}}_{\Theta}. \quad (5)$$

To ensure that Θ can be uniquely recovered from Y , certain conditions must be met. For example, it is clear that the total number of measurements cannot be smaller than the total number of unknowns, i.e., that we must have

$$\sum_{j=1}^J M_j \geq \dim(\Theta) = K_C(P) + \sum_{j=1}^J K_j(P). \quad (6)$$

However, only certain distributions of these measurements among the sensors will actually permit recovery. For example, the component $\theta_j \in \mathbb{R}^{K_j(P)}$ is measured only by sensor j , and so we require that

$$M_j \geq K_j(P) \quad (7)$$

for each $j \in \Lambda$. Taken together, conditions (6) and (7) state that each sensor must collect enough measurements to allow for recovery of the local innovation component, while the sensors collectively must acquire at least $K_C(P)$ extra measurements to permit recovery of the common component. While these conditions are indeed necessary for permitting recovery of Θ from Y (see Theorem 2), they are not sufficient—there are additional restrictions governing how these extra measurements may be allocated to permit recovery of the common component.

To appreciate the reason for these additional restrictions, consider the case where for some indices $n \in \{1, 2, \dots, N\}$ and $j \in \Lambda$, row n of P_C contains a 1 and row n of P_j contains a 1. Recalling the definitions of z_C and z_j from (2), this implies that both $z_C(n)$ and $z_j(n)$ have a corresponding entry in the unknown value vector Θ . In such an event, however, it is impossible to recover the values of both $z_C(n)$ and $z_j(n)$ from measurements of x_j alone because these pieces of information are added into the single element $x_j(n) = z_C(n) + z_j(n)$. Intuitively, since the correct value for $z_j(n)$ can only be inferred from y_j , it seems that the value $z_C(n)$ can only be inferred using measurements of other signals that do not feature the same overlap, i.e., from those $y_{j'}$ such that row n of $P_{j'}$ contains all zeros.

Based on the considerations above, we propose the following definition.

Definition 6: For a given location matrix P belonging to a common/innovation ESM \mathcal{P} and a given set of signals $\Gamma \subseteq \Lambda$, the *overlap size* $K_C(\Gamma, P)$ is the number of indices in which there is overlap between the common and innovation component supports at all signals $j \in \Gamma^C$:

$$K_C(\Gamma, P) := \left| \{n \in \{1, \dots, N\} : \text{row } n \text{ of } P_C \text{ is nonzero and } \forall j \in \Gamma^C, \text{ row } n \text{ of } P_j \text{ is nonzero}\} \right|. \quad (8)$$

We note that $K_C(\Lambda, P) = K_C(P)$ and $K_C(\emptyset, P) = 0$.

Relating to our discussion above, for each entry $n \in \{1, \dots, N\}$ counted in $K_C(\Gamma, P)$, we expect that some sensor in Γ must take one extra measurement to account for that entry of the common component—it is impossible to recover such entries from measurements made only by sensors outside Γ . Our first main result confirms that ensuring the sensors in every $\Gamma \subseteq \Lambda$ collectively acquire at least $K_C(\Gamma, P)$ extra measurements is indeed sufficient to permit recovery of Θ from Y .

Theorem 1: (Achievable, known P) Let X denote a signal ensemble, and let $P \in \mathcal{P}_F(X)$ be a full-rank location matrix in a common/innovation ESM \mathcal{P} . For each $j \in \Lambda$, let Φ_j be a random $M_j \times N$ matrix populated with i.i.d. Gaussian entries. If

$$\sum_{j \in \Gamma} M_j \geq \left(\sum_{j \in \Gamma} K_j(P) \right) + K_C(\Gamma, P) \quad (9)$$

for all subsets $\Gamma \subseteq \Lambda$, then with probability one over $\{\Phi_j\}_{j \in \Lambda}$, there exists a unique solution $\hat{\Theta}$ to the system of equations $Y = \Phi P \hat{\Theta}$, and hence, letting $\hat{X} := P \hat{\Theta}$ we have $\hat{X} = X$.

Our proof of Theorem 1 is presented in Section IV. The proof is based on a bipartite graph formulation that represents the dependencies between the obtained measurements Y and the coefficients in the value vector Θ . Intuitively, the bipartite graph arises from an interpretation of the matrix $\Upsilon = \Phi P$ as a biadjacency matrix [22]. The graph is fundamental both in the derivation of the number of measurements needed for each sensor and in the formulation of a combinatorial recovery procedure for the case where P is unknown; we revisit that problem in Section III-C below.

Although Theorem 1 can be invoked with any feasible location matrix, it yields the most favorable bounds when invoked using a location matrix that contains just D columns. One implication of this theorem is that, when a location matrix $P \in \mathcal{P}_F(X)$ is known, reconstruction of a signal ensemble X can be achieved using fewer than $\|x_j\|_0$ measurements at some or all of the sensors j . This highlights the benefit of joint reconstruction in DCS.

Our second main result establishes that the measurement bound presented in Theorem 1 cannot be improved. We defer the proof of the following theorem to Appendix A.

Theorem 2: (Converse) Let X denote a signal ensemble, and let $P \in \mathcal{P}_F(X)$ be a full-rank location matrix in a common/innovation ESM \mathcal{P} . For each $j \in \Lambda$, let Φ_j be an $M_j \times N$ matrix (not necessarily random). If

$$\sum_{j \in \Gamma} M_j < \left(\sum_{j \in \Gamma} K_j(P) \right) + K_C(\Gamma, P) \quad (10)$$

for some nonempty subset $\Gamma \subseteq \Lambda$, then there exists a value vector $\hat{\Theta}$ such that $Y = \Phi P \hat{\Theta}$ but $\hat{X} := P \hat{\Theta} \neq X$.

Example 2: Consider again the signal ensemble presented in Example 1. For the matrix P specified in (3), the overlap sizes are $K_C(\{1\}, P) = 0$ (since there is no overlap between common and innovation components in sensor 2), $K_C(\{2\}, P) = 1$ (since there is overlap in the common and innovation components at sensor 1 for index 1), and $K_C(\{1, 2\}, P) = K_C(P) = 2$. Alternatively, for the matrix \tilde{P} specified in (4), the overlap sizes are $K_C(\{1\}, \tilde{P}) = 1$, $K_C(\{2\}, \tilde{P}) = 0$, and $K_C(\{1, 2\}, \tilde{P}) = K_C(\tilde{P}) = 2$. Thus, for a decoder with knowledge of either one of these location matrices, Theorem 1 tells us that X can be uniquely recovered if $M_1 \geq 1$, $M_2 \geq 1$, and

$M_1 + M_2 \geq 3$. Conversely, Theorem 2 tells us that X cannot be uniquely recovered using either of these location matrices if $M_1 = 0$, if $M_2 = 0$, or if $M_1 = M_2 = 1$.

C. Identification of a Feasible Location Matrix

In general, when presented with only the measurements Y , it may be necessary for a decoder to find *both* a feasible location matrix $P \in \mathcal{P}_F(X)$ and a value vector Θ such that $X = P\Theta$. Just as identifying the sparse coefficient locations in single-signal CS can require more measurements than solving for the values if the locations are known [3], the multi-signal problem of jointly recovering P and Θ could require more measurements than specified in Theorem 1 for the case where P is known. Our final main result, however, guarantees that a moderate increase in the number of measurements beyond the bound specified in (9) is sufficient. The following is proved in Appendix B.

Theorem 3: (Achievable, unknown P) Let X denote a signal ensemble, and let \mathcal{P} denote a common/innovation ESM. For each $j \in \Lambda$, let Φ_j be a random $M_j \times N$ matrix populated with i.i.d. Gaussian entries. If there exists a full-rank location matrix $P^* \in \mathcal{P}_F(X)$ such that

$$\sum_{j \in \Gamma} M_j \geq \left(\sum_{j \in \Gamma} K_j(P^*) \right) + K_C(\Gamma, P^*) + |\Gamma| \quad (11)$$

for all subsets $\Gamma \subseteq \Lambda$, then X can be recovered from Y .

The achievable measurement bound in (11) can be met by taking just one additional measurement per sensor above the rate specified in (9); comparing with the converse bound in (10), we see that this is virtually as tight as possible. Like Theorem 1, Theorem 3 yields the most favorable bounds when invoked using a location matrix that contains just D columns.

The proof of Theorem 3 involves an algorithm based on an enumerative search over all $P \in \mathcal{P}$; this is akin to the ℓ_0 minimization problem in single-signal CS. Indeed, removing the common component and taking $J = 1$, our bound reduces to the classical single-signal CS result that $K + 1$ Gaussian random measurements suffice with probability one to enable recovery of a fixed K -sparse signal via ℓ_0 minimization [3, 23]. Although such an algorithm may not be practically implementable or robust to measurement noise, we believe that our Theorem 3 (taken together with Theorems 1 and 2) provides a theoretical foundation for understanding the core issues surrounding the measurement and reconstruction of signal ensembles in the context of ESMs.

Example 3: We once again revisit the signal ensemble presented in Example 1. Using either of the feasible matrices P specified in (3) or (4) for the purpose of evaluating the bound (11), Theorem 3 tells us that an a priori unknown feasible location matrix and corresponding value vector can be found to allow perfect recovery of the signal ensemble X , as long as $M_1 \geq 2$, $M_2 \geq 2$, and $M_1 + M_2 \geq 5$. For example, x_1 and x_2 can be recovered when $M_1 = 3$ and $M_2 = 2$. For this choice of M_1 and M_2 and for the location matrix P specified in (3), Figure 1 shows that there exists a matching that associates each element of the value vector Θ to a unique measurement. Our exposition of the graph based formulation (see Section IV) explains how the existence of such a matching

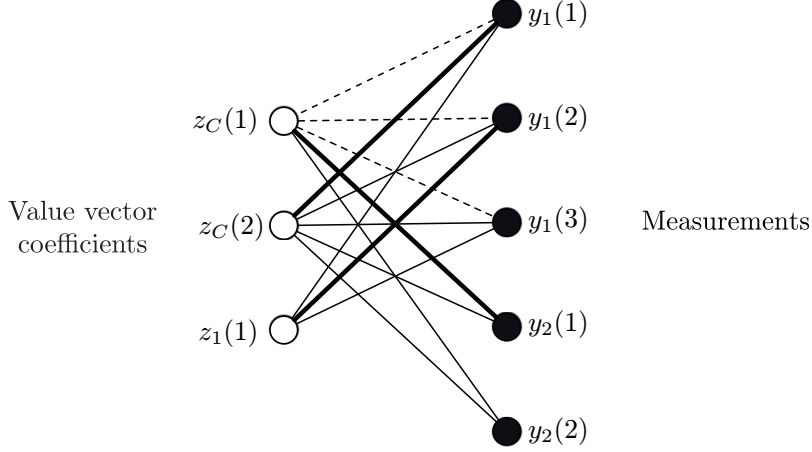


Fig. 1. Graphical representation of the dependencies between value vector coefficients and compressive measurements for the signal ensemble X and location matrix P discussed in Example 3. Each edge in this graph denotes a dependency of a measurement on a value vector coefficient, but dashed lines indicate dependencies that cannot be exploited due to overlap of common and innovation coefficients. Among the edges that remain, the thick solid lines indicate the existence of a matching from each value vector coefficient to a distinct measurement; the existence of such a matching ensures that the system of equations $Y = \Phi P \Theta$ is invertible (see Theorem 1 and its proof in Section IV). Measurements that remain unassigned in this matching can then be used to verify the correctness of the solution (see Theorem 3 and its proof in Appendix B).

ensures perfect recovery of Θ , given P , and our proof of Theorem 3 (see Appendix B) explains how the remaining measurements can be used to identify a feasible location matrix.

IV. CENTRAL PROOF AND BIPARTITE GRAPH FORMULATION

This section is dedicated to proving Theorem 1. In order to prove this theorem, we introduce a bipartite graph formulation that represents the dependencies between the obtained measurements Y and the coefficients in the value vector Θ .

A. Proof of Theorem 1

For brevity, we denote $K_C(P)$ and $K_j(P)$ simply as K_C and K_j . We denote the number of columns of P by

$$D' = D'(P) := K_C + \sum_{j \in \Lambda} K_j, \quad (12)$$

and note that $D' \geq D$. Because $P \in \mathcal{P}_F(X)$, there exists $\Theta \in \mathbb{R}^{D'}$ such that $X = P\Theta$. Because $Y = \Phi X$, Θ is a solution to $Y = \Phi P \Theta$.

We will argue that, with probability one over Φ , $\Upsilon := \Phi P$ has rank D' , and thus Θ is the unique solution to the equation $Y = \Phi P \Theta = \Upsilon \Theta$. To prove that Υ has rank D' , we invoke the following lemma, which we prove in Section IV-B.

Lemma 1: If (9) holds, then there exists a mapping $\mathcal{C} : \{1, 2, \dots, K_C\} \rightarrow \Lambda$, assigning each element of the common component to one of the sensors, such that for all nonempty subsets $\Gamma \subseteq \Lambda$,

$$\sum_{j \in \Gamma} M_j \geq \sum_{j \in \Gamma} (K_j + C_j), \quad (13)$$

where $C_j := |\{k \in \{1, 2, \dots, K_C\} : \mathcal{C}(k) = j\}|$ for each $j \in \Lambda$, and such that for each $k \in \{1, 2, \dots, K_C\}$, the k^{th} column of P_C is not a column of $P_{\mathcal{C}(k)}$.

Intuitively, the existence of such a mapping suggests that (i) each sensor has taken enough measurements to cover its own innovation component (requiring K_j measurements) and perhaps some of the common component, (ii) for any $\Gamma \subseteq \Lambda$, the sensors in Γ have collectively taken enough extra measurements to cover the requisite $K_C(\Gamma, P)$ elements of the common component, and (iii) the extra measurements are taken at sensors where the common and innovation components do not overlap. Formally, we will use the existence of such a mapping to prove that Υ has rank D' .

We proceed by noting that Υ has the block structure illustrated in (5), where each $\Phi_j P_C$ (respectively, $\Phi_j P_j$) is an $M_j \times K_C$ (respectively, $M_j \times K_j$) submatrix of Φ_j obtained by selecting columns from Φ_j according to the columns contained in P_C (respectively, P_j). Referring to (12), we see that, in total, Υ has D' columns. To argue that Υ has rank D' , we will consider a sequence of three matrices Υ_0 , Υ_1 , and Υ_2 constructed from modifications to Υ .

Construction of Υ_0 : We begin by letting Υ_0 denote the “partially zeroed” matrix obtained from Υ using the following construction:

- 1) Let $\Upsilon_0 = \Upsilon$ and $k = 1$.
- 2) For each j such that P_j has a column that matches column k of P_C (note that by Lemma 1 this cannot happen if $\mathcal{C}(k) = j$), let k' represent the column index of the full matrix P where this column of P_j occurs. Subtract column k' of Υ_0 from column k of Υ_0 . This forces to zero all entries of Υ_0 formerly corresponding to column k of the block $\Phi_j P_C$.
- 3) If $k < K_C$, then increment k and go to step 2.

The matrix Υ_0 is identical to Υ everywhere except on the first K_C columns, where any portion of a column equal to a column of $\Phi_j P_j$ to its right has been set to zero.³ Thus, Υ_0 satisfies the next two properties, which will be inherited by matrices Υ_1 and Υ_2 that we subsequently define:

- P1. Each entry of Υ_0 is either zero or a Gaussian random variable.
- P2. All Gaussian random variables in Υ_0 are i.i.d.

Finally, because Υ_0 was constructed only by subtracting columns of Υ from one another, $\text{rank}(\Upsilon_0) = \text{rank}(\Upsilon)$.

Construction of Υ_1 : We now let Υ_1 be the matrix obtained from Υ_0 using the following construction: For each $j \in \Lambda$, we select $K_j + C_j$ arbitrary rows from the portion of Υ_0 corresponding to sensor j (the first M_1 rows of

³We later show that with probability one, none of the columns become entirely zero.

Υ_0 correspond to sensor 1, the following M_2 rows correspond to sensor 2, and so on). The resulting matrix Υ_1 has

$$\sum_{j \in \Lambda} (K_j + C_j) = \left(\sum_{j \in \Lambda} K_j \right) + K_C = D'$$

rows; note that this is fewer than the number of rows in Υ_0 if $K_j + C_j < M_j$ for any j . Also, because Υ_1 was obtained by selecting a subset of rows from Υ_0 , it has D' columns (just like Υ_0) and satisfies $\text{rank}(\Upsilon_1) \leq \text{rank}(\Upsilon_0) = \text{rank}(\Upsilon)$.

Construction of Υ_2 : We now let Υ_2 be the $D' \times D'$ matrix obtained by permuting columns of Υ_1 using the following construction:

- 1) Let $\Upsilon_2 = []$, and let $j = 1$.
- 2) For each k such that $\mathcal{C}(k) = j$, let $\Upsilon_1(k)$ denote the k^{th} column of Υ_1 , and concatenate $\Upsilon_1(k)$ to Υ_2 , i.e., let $\Upsilon_2 \leftarrow [\Upsilon_2 \ \Upsilon_1(k)]$. There are C_j such columns.
- 3) Let $\Upsilon_{1,j}$ denote the columns of Υ_1 corresponding to the entries of $\Phi_j P_j$ (the innovation components of sensor j), and concatenate $\Upsilon_{1,j}$ to Υ_2 , i.e., let $\Upsilon_2 \leftarrow [\Upsilon_2 \ \Upsilon_{1,j}]$. There are K_j such columns.
- 4) If $j < J$, then increment j and go to Step 2.

In total, Step 2 chooses $\sum_{j=1}^J C_j = K_C$ columns, while Step 3 chooses $\sum_{j=1}^J K_j$ columns, and thus referring to (12), Υ_2 has $K_C + \sum_{j=1}^J K_j = D'$ columns. The number of rows is the same as that of Υ_1 , making Υ_2 a square matrix. Because Υ_1 and Υ_2 share the same columns up to reordering, it follows that

$$\text{rank}(\Upsilon_2) = \text{rank}(\Upsilon_1) \leq \text{rank}(\Upsilon). \quad (14)$$

Based on its dependency on Υ_0 , and following from Lemma 1, Υ_2 meets properties P1 and P2 defined above in addition to a third property:

P3. All entries along the main diagonal of Υ_2 are Gaussian random variables (none are deterministically zero). Property P3 follows because each diagonal element of Υ_2 will either be an entry of some $\Phi_j P_j$, which remains Gaussian throughout our constructions, or it will be an entry of some k^{th} column of some $\Phi_j P_C$ for which $\mathcal{C}(k) = j$. In the latter case, we know by Lemma 1 and the construction of Υ_0 (Step 2) that the k^{th} column of $\Phi_j P_C$ is not zeroed out, and thus the corresponding diagonal entry remains Gaussian throughout our constructions.

Having identified these three properties satisfied by Υ_2 , we will prove by induction that, with probability one over Φ , such a matrix has full rank.

Lemma 2: Let $\Upsilon^{(d-1)}$ be a $(d-1) \times (d-1)$ matrix having full rank. Construct a $d \times d$ matrix $\Upsilon^{(d)}$ as follows:

$$\Upsilon^{(d)} := \begin{bmatrix} \Upsilon^{(d-1)} & v_1 \\ v_2^T & \omega \end{bmatrix}$$

where $v_1, v_2 \in \mathbb{R}^{d-1}$ are column vectors with each entry being either zero or a Gaussian random variable, ω is a Gaussian random variable, and all random variables are i.i.d. and independent of $\Upsilon^{(d-1)}$. Then with probability one, $\Upsilon^{(d)}$ has full rank.

Applying Lemma 2 inductively D' times, the success probability remains one. It follows that with probability one over Φ , $\text{rank}(\Upsilon_2) = D'$. Combining this last result with (14), we conclude that $\text{rank}(\Upsilon) = D'$ with probability one over Φ . It remains to prove Lemma 2.

Proof of Lemma 2: When $d = 1$, $\Upsilon^{(d)} = [\omega]$, which has full rank if and only if $\omega \neq 0$, which occurs with probability one.

When $d > 1$, using expansion by minors, the determinant of $\Upsilon^{(d)}$ satisfies

$$\det(\Upsilon^{(d)}) = \omega \cdot \det(\Upsilon^{(d-1)}) + C,$$

where $C = C(\Upsilon^{(d-1)}, v_1, v_2)$ is independent of ω . The matrix $\Upsilon^{(d)}$ has full rank if and only if $\det(\Upsilon^{(d)}) \neq 0$, which is satisfied if and only if

$$\omega \neq \frac{-C}{\det(\Upsilon^{(d-1)})}.$$

By the inductive assumption, $\det(\Upsilon^{(d-1)}) \neq 0$ and ω is a Gaussian random variable that is independent of C and $\det(\Upsilon^{(d-1)})$. Thus, $\omega \neq \frac{-C}{\det(\Upsilon^{(d-1)})}$ with probability one. \square

B. Proof of Lemma 1

To prove Lemma 1, we apply tools from graph theory.

We introduce a bipartite graph $G = (V_V, V_M, E)$ that captures the dependencies between the entries of the value vector $\Theta \in \mathbb{R}^{D'}$ and the entries of the measurement vector $Y = \Phi P \Theta$. This graph is defined as follows. The set of *value vertices* V_V has elements with indices $d \in \{1, \dots, D'\}$ representing the entries $\Theta(d)$ of the value vector. The set of *measurement vertices* V_M has elements with indices (j, m) representing the measurements $y_j(m)$, with $j \in \Lambda$ and $m \in \{1, \dots, M_j\}$ (the range of possible m varies depending on j). The cardinalities for these sets are $|V_V| = D'$ and $|V_M| = \sum_{j \in \Lambda} M_j$. Finally, the set of edges E is defined according to the following rules:

- For every $d \in \{1, 2, \dots, K_C\} \subseteq V_V$ and $j \in \Lambda$ such that column d of P_C does not also appear as a column of P_j , we have an edge connecting d to each vertex $(j, m) \in V_M$ for $1 \leq m \leq M_j$.
- For every $d \in \{K_C + 1, K_C + 2, \dots, D'\} \subseteq V_V$, we consider the sensor j associated with column d of P , and we have an edge connecting d to each vertex $(j, m) \in V_M$ for $1 \leq m \leq M_j$.

An example graph for a distributed sensing setting appears in Figure 2.

We seek a *matching* within the bipartite graph $G = (V_V, V_M, E)$, namely, a subgraph (V_V, V_M, \bar{E}) with $\bar{E} \subseteq E$ that pairs each element of V_V with a unique element of V_M . Such a matching will immediately give us the desired mapping \mathcal{C} as follows: for each $k \in \{1, 2, \dots, K_C\} \subseteq V_V$, let $(j, m) \in V_M$ denote the single vertex matched to k by an edge in \bar{E} ; we then set $\mathcal{C}(k) = j$.

To prove the existence of such a matching within the graph, we invoke a version of Hall's marriage theorem for bipartite graphs [24]. Hall's theorem states that within a bipartite graph (V_1, V_2, E) , there exists a matching that assigns each element of V_1 to a unique element of V_2 if for any collection of elements $\Pi \subseteq V_1$, the set $E(\Pi)$ of neighbors of Π in V_2 has cardinality $|E(\Pi)| \geq |\Pi|$. To apply Hall's theorem in the context of our lemma, we will

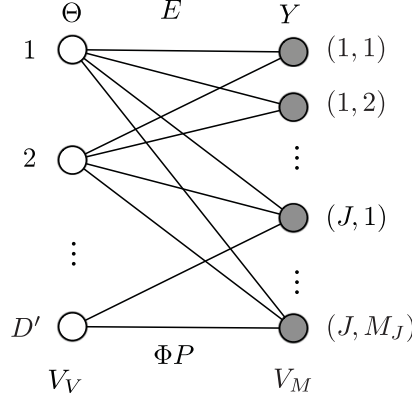


Fig. 2. The bipartite graph $G = (V_V, V_M, E)$ indicates the relationship between the value vector coefficients and the measurements.

show that if (9) is satisfied, then for any set $\Pi \subseteq V_V$ of entries in the value vector, the set $E(\Pi)$ of neighbors of Π in V_M has size $|E(\Pi)| \geq |\Pi|$.

Let us consider an arbitrary set $\Pi \subseteq V_V$. We let $S_\Pi = \{j \in \Lambda : (j, m) \in E(\Pi) \text{ for some } m\} \subseteq \Lambda$ denote the set of signal indices whose measurement vertices have edges that connect to Π . Since a connection between a value vertex and a measurement vertex at a given sensor implies a connection to all other measurement vertices for that sensor, it follows that $|E(\Pi)| = \sum_{j \in S_\Pi} M_j$. Thus, in order to satisfy Hall's condition for Π , we require

$$\sum_{j \in S_\Pi} M_j \geq |\Pi|. \quad (15)$$

We would now like to show that $\sum_{j \in S_\Pi} K_j + K_C(S_\Pi, P) \geq |\Pi|$, and thus if (9) is satisfied for all $\Gamma \subseteq \Lambda$, then (15) is satisfied in particular for $S_\Pi \subseteq \Lambda$.

In general, the set Π may contain vertices for both common components and innovation components. We write $\Pi = \Pi_C \cup \Pi_I$ to denote the disjoint union of these two sets.

By construction, $|\Pi_I| \leq \sum_{j \in S_\Pi} K_j$, because Π_I cannot include any innovation component outside the set of sensors S_Π . We will also argue that $|\Pi_C| \leq K_C(S_\Pi, P)$ as follows. By definition, for a set $\Gamma \subseteq \Lambda$, $K_C(\Gamma, P)$ counts the number of columns in P_C that also appear in P_j for all $j \notin \Gamma$. By construction, for each $k \in \Pi_C$, vertex k has no connection to vertices (j, m) for $j \notin S_\Pi$, and so it must follow that the k^{th} column of P_C is present in P_j for all $j \notin S_\Pi$. Thus, the index k is among the indices counted in the definition (8) of $K_C(S_\Pi, P)$, and therefore $|\Pi_C| \leq K_C(S_\Pi, P)$.

We conclude that $|\Pi| = |\Pi_I| + |\Pi_C| \leq \sum_{j \in S_\Pi} K_j + K_C(S_\Pi, P)$, and so (9) implies (15) for any Π , and so Hall's condition is satisfied, and a matching exists. Finally, consider any set $\Gamma \subseteq \Lambda$. To confirm that (13) holds for this set, note that there are a total of $\sum_{j \in \Gamma} M_j$ vertices $(j, m) \in V_M$ such that $j \in \Gamma$. Each of these vertices is matched to at most one vertex in V_V , which must correspond either to an innovation component counted in K_j for some $j \in \Gamma$ or to a common component indexed by some k such that $\mathcal{C}(k) \in \Gamma$. It follows that $\sum_{j \in \Gamma} M_j \geq \sum_{j \in \Gamma} (K_j + C_j)$.

□

V. DISCUSSION

In this paper, we have introduced the ensemble sparsity model (ESM) framework for modeling intra- and inter-signal correlations within a collection of sparse signals. This framework is based on a factored representation of the signal ensemble that decouples its location information from its value information. We have also proposed an analytical framework based on bipartite graphs that allowed us, in the context of a common/innovation ESM \mathcal{P} , to characterize the numbers of measurements M_1, M_2, \dots, M_J needed for successful recovery of a signal ensemble X . Our bounds highlight the benefit of joint reconstruction in distributed compressive sensing (DCS), since sparse signals can be recovered from fewer measurements than their nominal sparsity level would indicate.

The factored representation that we have proposed for modeling sparse signal ensembles is closely related to the recently proposed union-of-subspaces modeling frameworks for CS [25–28]. What is particularly novel about our treatment is the explicit consideration of the block structure of matrices such as P and Φ , and the explicit accounting for measurement bounds on a sensor-by-sensor basis. Most of the conventional union-of-subspaces theory in CS is intended to characterize the number of measurements required to recover a vector X from measurements ΦX , where Φ is a dense matrix.

Finally, as we have discussed in Section II-E, common/innovation ESMs can be populated using various choices of matrices P of the form (1). Our bounds in Section III are relatively agnostic to such design choices. Past experience, however, has indicated that practical algorithms for signal recovery can benefit from being tuned to the particular type of signal correlations under consideration [3, 11–14, 18–21, 29–33]. Our focus in this paper has been not on tractable recovery algorithms or robustness issues, but rather on foundational limits governing how the measurements may be amortized across the sensors while preserving the information required to uniquely identify the sparse signal ensemble. However, we believe that our results and our analytical framework may pave the way for a better, perhaps more unified, development of practical DCS algorithms.

APPENDIX A

PROOF OF THEOREM 2

As in (12), we let D' denote the number of columns in P . Because $P \in \mathcal{P}_F(X)$, there exists $\Theta \in \mathbb{R}^{D'}$ such that $X = P\Theta$. Because $Y = \Phi X$, then Θ is a solution to $Y = \Phi P\Theta$. We will argue for $\Upsilon := \Phi P$ that $\text{rank}(\Upsilon) < D'$, and thus there exists $\hat{\Theta} \neq \Theta$ such that $Y = \Upsilon\Theta = \Upsilon\hat{\Theta}$. Since P has full rank, it follows that $\hat{X} := P\hat{\Theta} \neq P\Theta = X$.

We let Υ_0 be the “partially zeroed” matrix obtained from Υ using the identical procedure detailed in Section IV-A. Again, because Υ_0 was constructed only by subtracting columns of Υ from one another, it follows that $\text{rank}(\Upsilon_0) = \text{rank}(\Upsilon)$.

Suppose that $\Gamma \subseteq \Lambda$ is a set for which (10) holds. We let Υ_3 be the submatrix of Υ_0 obtained by selecting the following columns:

- For any $k \in \{1, 2, \dots, K_C\}$ such that column k of P_C also appears as a column in P_j for all $j \notin \Gamma$, we include column k of Υ_0 as a column in Υ_3 . There are $K_C(\Gamma, P)$ such columns k .

- For any $k \in \{K_C + 1, K_C + 2, \dots, D'\}$ such that column k of P corresponds to an innovation for some sensor $j \in \Gamma$, we include column k of Υ_0 as a column in Υ_3 . There are $\sum_{j \in \Gamma} K_j$ such columns k .

This submatrix has $\sum_{j \in \Gamma} K_j + K_C(\Gamma, P)$ columns. Because Υ_0 has the same size as Υ (see Section IV-A), and in particular has only D' columns, then in order to have that $\text{rank}(\Upsilon_0) = D'$, it is necessary that all $\sum_{j \in \Gamma} K_j + K_C(\Gamma, P)$ columns of Υ_3 be linearly independent.

Based on the method described for constructing Υ_0 , it follows that Υ_3 is zero for all measurement rows not corresponding to the set Γ . These rows were nonzero only for two sets of columns of Υ_0 : (i) the columns corresponding to the innovations for signals $j \notin \Gamma$, and (ii) the columns $k \in \{1, 2, \dots, K_C\}$ for which the k^{th} column of P_C appears in none of the matrices $P_j, j \notin \Gamma$. Both of these sets of columns are discarded during the construction of Υ_3 . Therefore, consider the submatrix Υ_4 of Υ_3 obtained by selecting only the measurement rows corresponding to the set Γ . Because all the rows discarded from Υ_3 are zero, it follows that $\text{rank}(\Upsilon_3) = \text{rank}(\Upsilon_4)$. However, since Υ_4 has only $\sum_{j \in \Gamma} M_j$ rows, we invoke (10) and have that $\text{rank}(\Upsilon_3) = \text{rank}(\Upsilon_4) \leq \sum_{j \in \Gamma} M_j < \sum_{j \in \Gamma} K_j + K_C(\Gamma, P)$. Thus, all $\sum_{j \in \Gamma} K_j + K_C(\Gamma, P)$ columns of Υ_3 cannot be linearly independent, and so Υ does not have full rank. This means that there exists $\hat{\Theta} \neq \Theta$ such that $Y = \Upsilon\Theta = \Upsilon\hat{\Theta}$, and thus we cannot distinguish between the two solutions $\hat{X} := P\hat{\Theta} \neq P\Theta = X$. \square

APPENDIX B

PROOF OF THEOREM 3

Given the measurements Y and measurement matrix Φ , we will show that it is possible to recover some $P \in \mathcal{P}_F(X)$ and a corresponding vector Θ such that $X = P\Theta$ using the following algorithm.

- Extract from each measurement vector y_j its final entry, and sum these entries to obtain the quantity $\bar{y} = \sum_{j \in \Gamma} y_j(M_j)$. Similarly, add the corresponding rows of Φ into a single row $\bar{\phi}^T$. The row vector $\bar{\phi}^T$ is a concatenation of the final rows of the matrices Φ_j , and thus its entries are i.i.d. Gaussian. Note that $\bar{y} = \bar{\phi}^T X$; this quantity will be used in a cross-validation step below.
- Group the remaining $\left(\sum_{j \in \Lambda} M_j\right) - J$ measurements into a vector \bar{Y} , and let $\bar{\Phi}$ contain the corresponding rows of Φ . We note that $\bar{\phi}$ is independent from $\bar{\Phi}$ and that $\bar{Y} = \bar{\Phi}X$.
- For each matrix $P \in \mathcal{P}$ such that $\bar{Y} \in \text{colspan}(\bar{\Phi}P)$, choose a single solution Θ_P to $\bar{Y} = \bar{\Phi}P\Theta_P$ independently of $\bar{\phi}$. Then, perform the following cross-validation: if $\bar{y} = \bar{\phi}^T P\Theta_P$, then return the estimate $\hat{X} = P\Theta_P$; otherwise, continue with the next matrix P .

We begin by noting that there exists at least one matrix $P \in \mathcal{P}$ for which $\bar{Y} \in \text{colspan}(\bar{\Phi}P)$ and for which $X = P\Theta_P$. In particular, consider the matrix $P^* \in \mathcal{P}_F(X)$ mentioned in the theorem statement. Because (11) holds for P^* , Theorem 1 guarantees that with probability one, $\bar{\Phi}P^*$ will have full rank, and so there is a unique solution Θ_{P^*} to $\bar{Y} = \bar{\Phi}P^*\Theta_{P^*}$. Since $P^* \in \mathcal{P}_F(X)$ and P^* is full rank, we know that $X = P^*\Theta_{P^*}$. Also, since $Y = \Phi X$, we know that $\bar{y} = \bar{\phi}^T P^*\Theta_{P^*}$, and so this matrix will clear the cross-validation step.

Now suppose that, for some $P \in \mathcal{P}$, the algorithm above considers a candidate solution Θ_P to $\bar{Y} = \bar{\Phi}P\Theta_P$, but suppose also that $X \neq P\Theta_P$. The algorithm will fail to discard this incorrect solution if Θ_P passes the cross-

validation test, i.e., if $\bar{\phi}^T P\Theta_P = \bar{y} = \bar{\phi}^T X$. Recall, however, that $\bar{\phi}$ is an i.i.d. Gaussian random vector and that it is independent of both X and $P\Theta_P$. It then follows that $\bar{\phi}$ is orthogonal to $X - P\Theta_P$ with probability zero, and therefore we will have $\bar{\phi}^T (X - P\Theta_P) \neq 0$ (equivalently, $\bar{\phi}^T P\Theta_P \neq \bar{y}$) with probability one. Therefore, this incorrect solution will be discarded with probability one. Since \mathcal{P} contains only a finite number of matrices, the probability of cross-validation discarding all incorrect solutions remains one. \square

ACKNOWLEDGMENTS

Thanks to Emmanuel Candès, Albert Cohen, Ron DeVore, Anna Gilbert, Illya Hicks, Robert Nowak, Jared Tanner, and Joel Tropp for informative and inspiring conversations. Portions of this work were performed at Princeton University, the University of Michigan, and the Technion by MFD, MBW, and DB, respectively. The authors thank these institutions (and in particular Prof. Tsachy Weissman) for their hospitality.

REFERENCES

- [1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.
- [2] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [3] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing," Rice University, Houston, TX, Tech. Rep. TREE0612, Nov. 2006, available at <http://dsp.rice.edu/cs/>.
- [4] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
- [5] D. Needell and J. A. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.
- [6] W. U. Bajwa, J. D. Haupt, A. M. Sayeed, and R. D. Nowak, "Joint source-channel communication for distributed estimation in sensor networks," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3629–3653, Oct. 2007.
- [7] M. Rabbat, J. D. Haupt, A. Singh, and R. D. Nowak, "Decentralized compression and predistribution via randomized gossiping," in *Int. Workshop Info. Proc. Sensor Networks (IPSN)*, Nashville, TN, Apr. 2006.
- [8] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *Int. Workshop Info. Proc. Sensor Networks (IPSN)*, Cambridge, MA, Apr. 2007.
- [9] S. Aeron, M. Zhao, and V. Saligrama, "On sensing capacity of sensor networks for the class of linear observation, fixed SNR models," in *IEEE Workshop Stat. Signal Proc. (SSP)*, Madison, WI, Aug. 2007, pp. 388–392.
- [10] M. F. Duarte, S. Sarvotham, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Theoretical performance limits for jointly sparse signals via graphical models," Rice University, Houston, TX, Tech. Rep. TREE0809, July 2008.
- [11] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Allerton Conf. Comm., Control, Comput.*, Monticello, IL, Sept. 2005.
- [12] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Asilomar Conf. Signals, Systems and Computers*, Pacific Grove, CA, Nov. 2005.
- [13] M. B. Wakin, S. Sarvotham, M. F. Duarte, D. Baron, and R. G. Baraniuk, "Recovery of jointly sparse signals from few random projections," in *Proc. Neural Info. Proc. Sys. (NIPS)*, Vancouver, Canada, Nov. 2005.
- [14] M. F. Duarte, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Universal distributed sensing via random projections," in *Int. Workshop Info. Proc. Sensor Networks (IPSN)*, Nashville, TN, Apr. 2006.
- [15] M. F. Duarte, S. Sarvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Performance limits for jointly sparse signals via graphical models," in *Workshop Sensor, Signal, Info. Proc. (SENSIP)*, Sedona, AZ, May 2008.
- [16] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, July 1973.
- [17] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

- [18] A. Jalali, P. Ravikumar, S. Sanghavi, and C. Ruan, "A dirty model for multi-task learning," in *Proc. Neural Info. Proc. Sys. (NIPS)*, vol. 23, 2010.
- [19] M. Fornassier and H. Rauhut, "Recovery algorithms for vector valued data with joint sparsity constraints," *SIAM Journal on Numerical Analysis*, vol. 46, no. 2, pp. 577–613, 2008.
- [20] M. Mishali and Y. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Proc.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.
- [21] S. R. Schnelle, J. N. Laska, C. Hegde, M. F. Duarte, M. Davenport, and R. G. Baraniuk, "Texas Hold'Em algorithms for distributed compressive sensing," in *IEEE Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP)*, Dallas, TX, Mar. 2010.
- [22] C. Berge, *The Theory of Graphs*. Mineola, NY: Dover, 2001.
- [23] R. Venkataramani and Y. Bresler, "Further results on spectrum blind sampling of 2D signals," in *IEEE Int. Conf. Image Proc. (ICIP)*, vol. 2, Chicago, Oct. 1998.
- [24] D. B. West, *Introduction to Graph Theory*. Prentice Hall, 1996.
- [25] Y. M. Lu and M. N. Do, "Sampling signals from a union of subspaces," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 41–47, Mar. 2008.
- [26] T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1872–1882, Apr. 2009.
- [27] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.
- [28] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Inf. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.
- [29] Y. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.
- [30] E. van den Berg and M. P. Friedlander, "Joint-sparse recovery from multiple measurements," University of British Columbia, Vancouver, Canada, Tech. Rep. TR-2009-07, Apr. 2009.
- [31] K. Lee and Y. Bresler, "Subspace-augmented MUSIC for joint sparse recovery with any rank," Apr. 2010, preprint, arXiv:1004.3071 [cs.IT].
- [32] J. M. Kim, O. K. Lee, and J. C. Ye, "Compressive MUSIC: A missing link between compressive sensing and array signal processing," Apr. 2010, preprint, arXiv:1004.4398v2 [cs.IT].
- [33] M. Davies and Y. Eldar, "Rank awareness in joint sparse recovery," Apr. 2010, preprint, arXiv:1004.4529 [cs.IT].